

Stochasticité et épidémie de COVID-19

Thomas Bénéteau, Baptiste Elie, Samuel Alizon, Mircea T Sofonea

▶ **To cite this version:**

Thomas Bénéteau, Baptiste Elie, Samuel Alizon, Mircea T Sofonea. Stochasticité et épidémie de COVID-19. [Rapport de recherche] 8, Centre National de la Recherche Scientifique (CNRS); Institut de Recherche pour le Développement (IRD); Université de Montpellier (UM), FRA. 2020. hal-02950660

HAL Id: hal-02950660
https://hal.science/hal-02950660

Submitted on 30 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Préambule
Introduction
Le modèle stochastique
Résultats
Discussion
Sources et remerciements

Stochasticité et épidémie de COVID-19

Groupe de modélisation de l'équipe ETE (https://www.mivegec.ird.fr/fr/contact/160-francais/equipes/1205-ete) (Laboratoire MIVEGEC (http://mivegec.ird.fr), CNRS, IRD, Université de Montpellier)

Le 20 mai 2020 (mis à jour le 25 mai 2020)

Préambule

Ce rapport a été produit à des fins académiques et ne constitue pas un support de prise de décision. De plus, les calculs réalisés ici sont faits avec des hypothèses volontairement simplifiées.

L'ensemble de nos rapports, notes et analyses sur la pandémie de COVID-19 est disponible sur cette page (http://covid-ete.ouvaton.org).

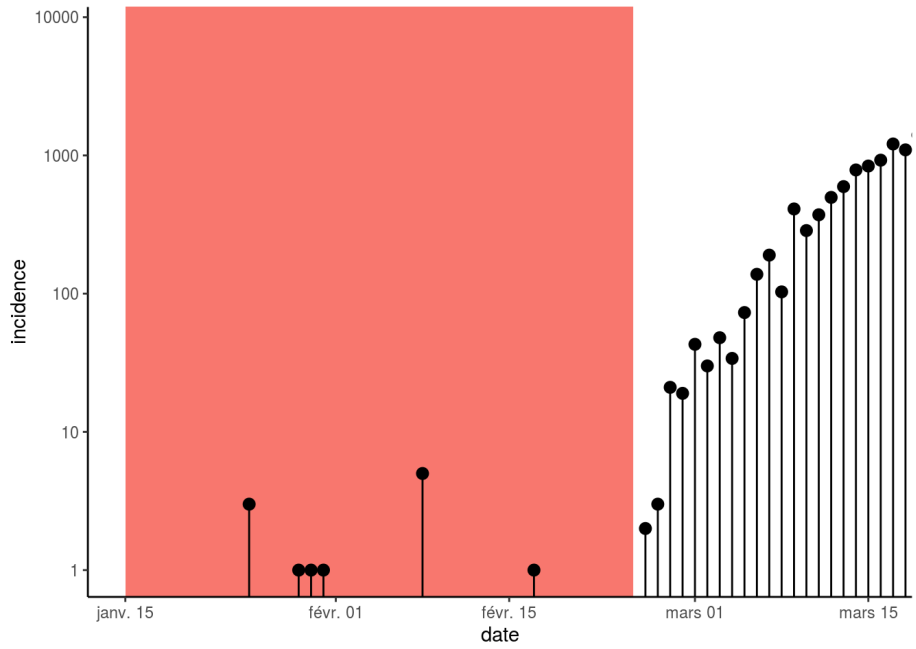
En matière de santé publique et pour toute question, nous recommandons de consulter et suivre les instructions officielles disponibles sur https://www.gouvernement.fr/info-coronavirus (https://www.gouvernement.fr/info-coronavirus)

L'Organisation Mondiale de la Santé (OMS) dispose aussi d'un site très complet https://www.who.int/fr/emergencies/diseases/novel-coronavirus-2019 (https://www.who.int/fr/emergencies/diseases/novel-coronavirus-2019)

Introduction

Une des raisons pour lesquelles les flambées épidémiques sont difficiles à prédire est que lorsque peu de personnes sont infectées, toutes les variations individuelles comptent. On regroupe ces variations imprévisibles sous le nom de **stochasticité** (ou hasard), du grec στοχαστικός, que l'on pourrait traduire par « habile à deviner ». Mathématiquement, le théorème qui explique que certains aspects des épidémies soient prédictibles est connu sous le nom de *loi des grands nombres*. Celle-ci stipule que le comportement moyen d'un grand nombre d'événements aléatoires analogues et indépendants converge vers l'espérance commune à ces événements.

Le graphique ci-dessous représente l'incidence quotidienne des dépistages de COVID-19 en France (attention l'axe des ordonnées est en échelle logarithmique). On distingue une première période où il se passe plusieurs jours sans que des cas soient détectés (en rouge) et une seconde période à partir du 27 février où la courbe d'incidence commence à croître linéairement sur le graphique (donc exponentiellement car l'échelle est logarithmique).



Au cours de la première période (en rouge), la dynamique est difficilement prévisible et on la caractérise comme étant stochastique. Une fois que la prévalence est élevée, un modèle de croissance capture bien la dynamique qui est dite déterministe.

Notre but principal est de déterminer de **dater le début de la vague épidémique** en France. Pour ce faire, nous allons analyser la partie initiale de l'épidémie à l'aide de modèles stochastiques et donc travailler avec des probabilités.

Attention, comme nous le détaillons dans notre Note sur les origines de l'épidémie (Note_origine.html), il faut bien distinguer la vague épidémique des premières introductions dans le pays. En effet, du fait des processus stochastiques, les chaînes de transmission peuvent s'éteindre toutes seules même si le nombre de reproduction moyen est supérieur à 1. *Il est donc possible de détecter des cas avant le début de la vague épidémique.*

De même, nous allons discuter de la date de fin de l'épidémie, qui est également un processus influencé par les effets stochastiques lorsque le nombre d'individus infectés devient faible. Nous nous placerons dans un scénario simpliste dans lequel le confinement est prolongé, et donc le nombre de reproduction (Note_R0.html) du virus est maintenu à son niveau le plus bas, afin de voir au bout de combien de jours on peut être sûrs à 95 % que le virus a disparu de la population.

Le modèle stochastique

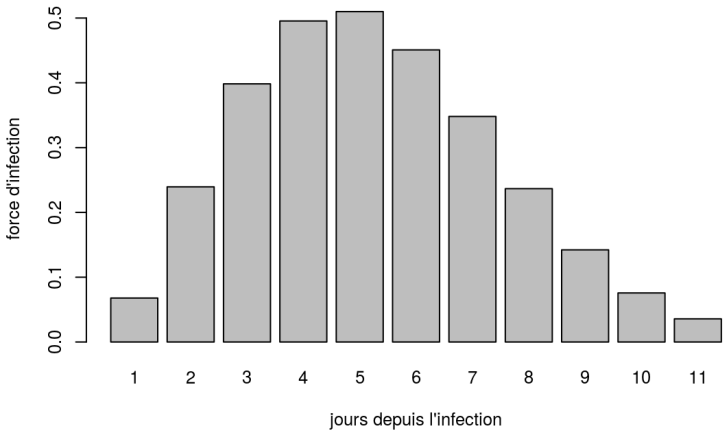
Notre modèle repose sur deux paramètres essentiels :

- le nombre de reproduction de base (Note_R0.html), $\mathcal{R}_0$, c'est-à-dire le nombre moyen d'infectés secondaires par cas, au début de l'épidémie,
- le temps de génération, c'est à dire la différence entre la date à laquelle une personne est infectée et la date à laquelle elle infecte une autre personne.

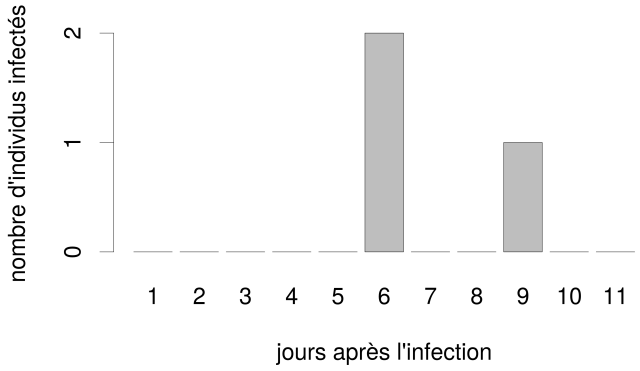
Le $\mathcal{R}_0$ peut être estimé à partir des données d'incidence de l'épidémie et des données d'intervalle de génération. Pour le virus du Covid-19, sa valeur est estimée à environ 3 (voir notre Note sur le sujet (http://covid-ete.ouvaton.org/Note_R0.html#en_france)).

Le temps de génération est en revanche beaucoup plus difficile à mesurer, car les dates de contaminations sont rarement connues, sinon entachées d'incertitudes. Lorsque l'on parvient à retracer l'histoire d'une infection entre individus, on remplace alors le temps de génération par l'*intervalle sériel*, c'est-à-dire la différence entre la date d'apparition des symptômes chez une personne et celle chez une personne qu'elle a infectée. À ce jour, les données qui permettraient de calculer cet intervalle en France ne sont pas accessibles et nous nous basons donc sur des distributions calculées en Asie, par exemple par Nishiura *et alii* (2020, *Int J Inf Dis* (https://linkinghub.elsevier.com/retrieve/pii/S1201971220301193)).

En combinant $\mathcal{R}_0$ avec la distribution du temps de génération (approché par l'intervalle sériel), on obtient la distribution du nombre moyen d'infections secondaires par jour post-contamination (on peut remarquer que cette distribution ne représente pas une loi de probabilités puisque elle somme à $\mathcal{R}_0$ et non à 1). Nous nous référerons à cette distribution comme étant la *force d'infection* individuelle et elle est représentée sur la figure suivante.



Sous l'hypothèse d'un taux de contacts homogène dans la population (notamment en absence de structure spatiale, une approximation connue sous le nom de *champ moyen*) et si l'infectiosité des individus est identique (il n'y a pas de super-propagateurs (https://fr.wikipedia.org/wiki/Superinfecteur), ou *superspreaders* en anglais), alors le nombre de personnes qu'une personne contamine un jour donné peut être modélisé par une loi de **Poisson** (https://fr.wikipedia.org/wiki/Loi_de_Poisson), dont le paramètre est sa force d'infection au jour considéré.



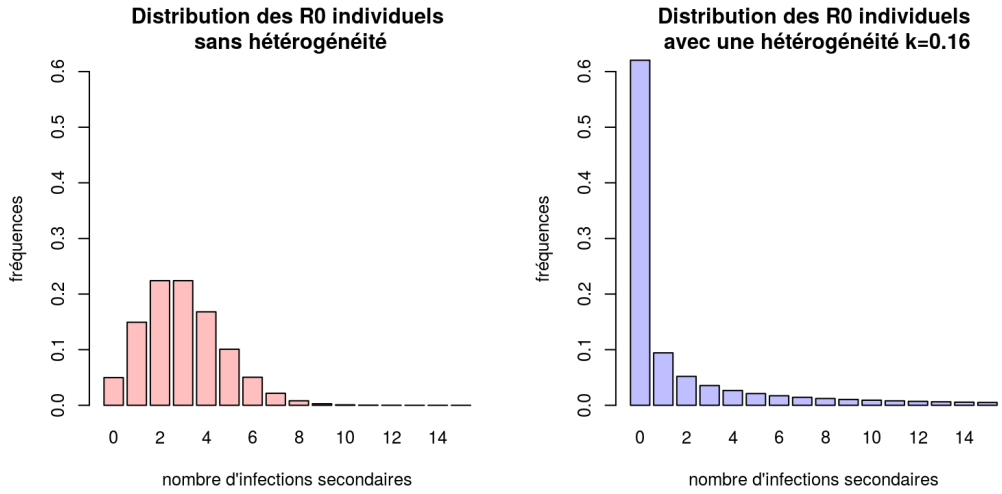
La figure ci-contre montre un exemple de la répartition des infections causées par un individu au cours de son infection: dans cet exemple, l'individu a infecté trois personnes : deux lors du 6^{e} jour de son infection, et une au 9^{e} jour.

En pratique, lors d'une épidémie, il y a toujours de l'hétérogénéité, que ce soit au niveau de la structure spatiale ou au niveau des individus car certains ont plus de contacts que d'autres. Afin de simuler cette hétérogénéité, nous faisons l'hypothèse que la force d'infection individuelle, qui jusque là était fixée à $\mathcal{R}_0$, suit une loi *gamma* (https://fr.wikipedia.org/wiki/Loi_binomiale_n%C3%A9gative) avec un paramètre de forme k et une moyenne $\mathcal{R}_0$. Comme cette force d'infection est le paramètre de loi de *Poisson* déterminant le nombre d'infections causées par un individu au jour suivant, on aboutit à ce que ce nombre suive une loi *binomiale négative* (https://fr.wikipedia.org/wiki/Loi_binomiale_n%C3%A9gative) de paramètre de dispersion k et de moyenne $\mathcal{R}_0$.

Plus le paramètre k est petit, plus l'hétérogénéité entre individus est importante. La loi de Poisson est en fait un cas limite de la loi binomiale négative, avec un paramètre de dispersion infini. Notre approche est similaire à celle utilisée par Lloyd-Smith *et alii* (2005, *Nature* (https://www.nature.com/articles/nature04153)), qui ont d'ailleurs estimé que pour le SRAS-Cov1 on avait $k = 0,16$. Par ailleurs, quelques études ont

estimé que la transmission du SARS-Cov2 était également assez hétérogène: une première étude, basée sur 391 cas en Chine, estimait un paramètre de dispersion $k = 0,58$ (Bi *et alii*, 2020 *MedRXiv* (https://www.medrxiv.org/content/10.1101/2020.03.03.20028423v3.full.pdf)). Cependant une seconde étude basée sur le nombre de cas importés face au nombre de cas endémiques en début d'épidémie, pour chaque pays du Monde, estimait une dispersion plus importante, avec un $k = 0,1$ (Endo *et alii*, 2020 *Wellcome Open Research* (https://d212y8ha88k086.cloudfront.net/manuscripts/17377/bbd75255-2e3a-4f30-96c9-28142e604f0f_15842_-_akira_endo.pdf?doi=10.12688/wellcomeopenres.15842.1&numberOfBrowsableCollections=5&numberOfBrowsableInstitutionalCollections=0&numberOfBrowsableGateways=13)).

La figure suivante représente la fréquence des $\mathcal{R}_0$ individuels observés dans une population sans hétérogénéité (avec la loi de *Poisson*) ou bien dans une population hétérogène (avec une loi *Binomiale Négative* de paramètre d'hétérogénéité $k = 0,16$). On peut voir qu'une majorité des infections n'engendre aucune infection secondaire lorsqu'on introduit de l'hétérogénéité, alors que quelques individus sont à l'origine d'un grand nombre d'infections (super-propagateurs).



En sommant les forces d'infections individuelles, on obtient le nombre de nouvelles infections causées chaque jour. Le schéma suivant explicite le calcul des infections que $y_{\text{inf},i}$ personnes infectées à un jour i causent les jours suivants.

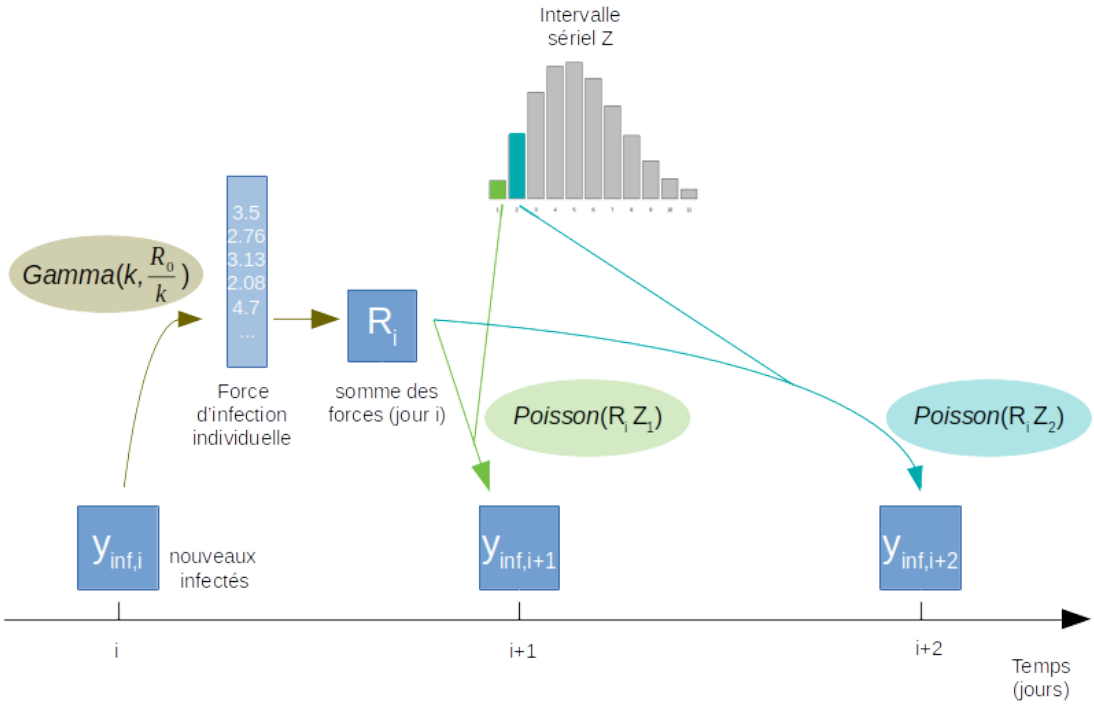


Schéma illustrant l'algorithme de génération du nombre de nouvelles infections quotidiennes avec de l'hétérogénéité individuelle. Sans hétérogénéité individuelle, l'étape du tirage au sort de la force d'infection individuelle est remplacée par une force d'infection fixe, valant ici $R_0 = 3$.

Résultats

Origine de la vague épidémique

Nous avons tout d'abord estimer la date du début de l'épidémie à partir des données de mortalité dans les hôpitaux français publiées par Santé Publique France (https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/). Ces séries temporelles sont parmi les caractérisations de l'épidémie les plus fiables dont nous disposons.

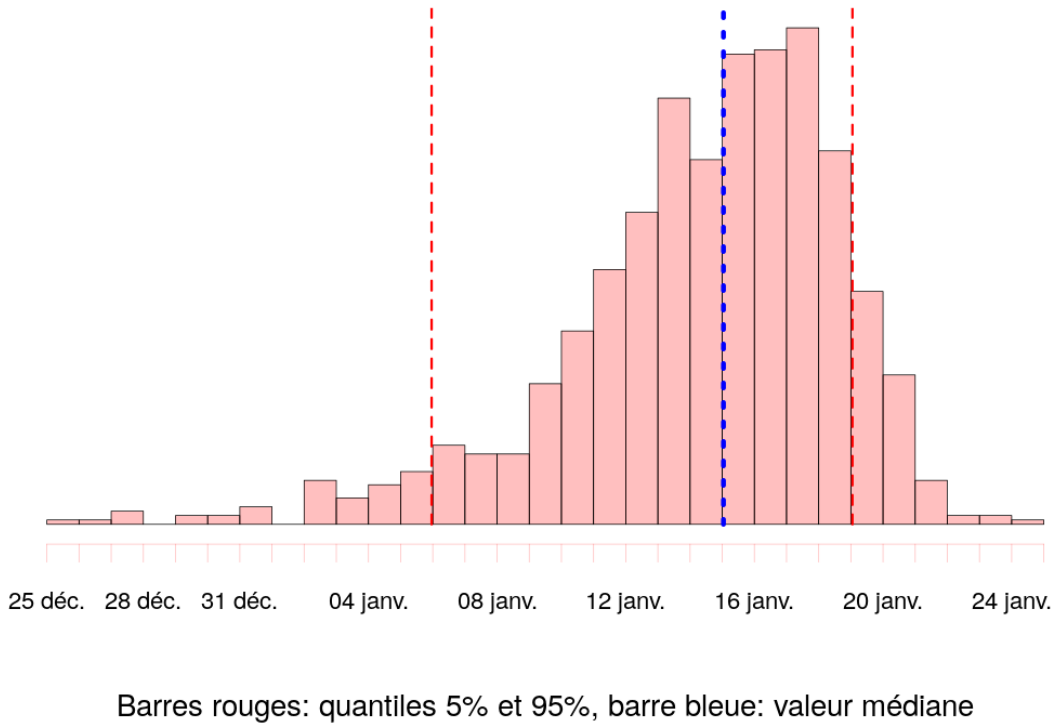
Afin d'estimer la date d'origine de la vague épidémique, nous avons fait les hypothèses suivantes :

- Un seul individu a initié la vague épidémique. Cette hypothèse n'est pas forcément réaliste (plusieurs chaînes de transmission ont pu contribuer à la vague), mais notre but est plus d'étudier l'impact des paramètres du modèle sur la date d'origine.

- La proportion de létalité (Note_IFR.html) de l'infection au SARS-Cov2 est de 0,5 %.
- Il existe un délai entre l'infection d'une personne et son décès. Suivant les données de Linton *et alii* (2020, *J Clin Med* (http://dx.doi.org/10.3390/jcm9020538)), nous avons estimé que ce délai suit une loi *log-normale* (https://fr.wikipedia.org/wiki/Loi_log-normale) de moyenne de 20,1 jours et d'écart-type 7,3 jours.

En pratique, nous avons réalisé 5.000 simulations stochastiques de flambées épidémiques et avons relevé le nombre de jours nécessaires pour que l'incidence dépasse les 100 décès quotidiens. Nous avons choisi ce seuil, atteint en rélité le 20 mars 2020, car lorsque le nombre d'infections quotidiennes aussi élevé, le seuil de flambée épidémique est dépassé et le rôle de la stochasticité dans la croissance de l'épidémie devient négligeable (Hartfield & Alizon 2014 *PLoS Path* (https://doi.org/10.1371/journal.ppat.1003277)). De plus, nous pouvons estimer que les effets du confinement ne se font pas encore ressentir sur l'évolution du nombre de décès quotidiens, quatre jours après son intiation, ce qui nous permet de maintenir un $\mathcal{R}(t)$ constant durant toute la période.

En retirant le nombre de jours nécessaires pour arrivé au seuil fixé dans la simulation à la véritable date à laquelle ce seuil a été atteint, on obtient une estimation de la date de début de l'épidémie en fonction des hypothèses du modèle.



Distribution des dates d'origine de la vague épidémique pour atteindre 100 décès le 20 mars

Au final, avec le modèle sans hétérogénéité individuelle, on obtient une valeur de début de la vague épidémique médiane au **15 janvier 2020**, et 95 % des valeurs comprises entre le 2 et le 19 janvier.

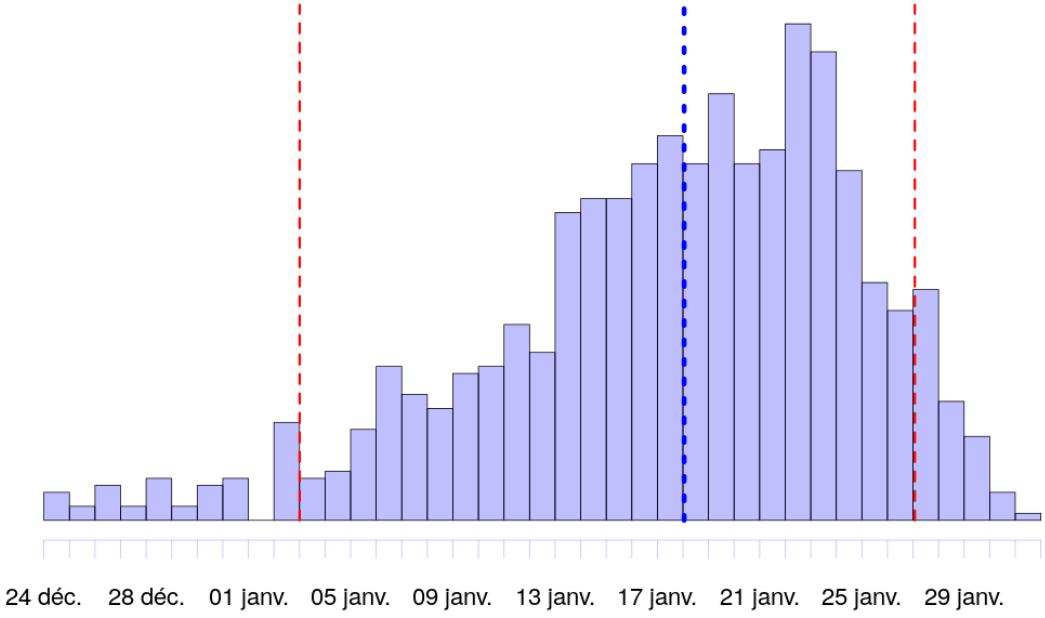
À partir de ces simulations, on peut aussi calculer la probabilité d'extinction d'une chaîne de transmission initiée par un individu. Pour nos hypothèses, notamment le $\mathcal{R}_0$ fixé à 3, cette probabilité est de 7 %. Ceci est cohérent avec les résultats analytiques obtenus sur des modèles de naissance et de mort sans mémoire.

Impact de l'hétérogénéité

L'hétérogénéité entre les individus, en particulier les évènements de super-propagation, sont connus pour affecter la dynamique d'une épidémie. Plus cette hétérogénéité est importante, plus l'épidémie a des chances de s'éteindre toute seule comme on le détaille dans cet article (https://www.larecherche.fr/covid-19-coronavirus-epid%C3%A9miologie/les-hasards-variables-des-%C3%A9mergences-virales). *A contrario*, les épidémies qui échappent à l'extinction stochastique croissent alors plus rapidement en moyenne (Lloyd-Smith *et alii* 2005 *Nature* (https://www.nature.com/articles/nature04153)).

Dans le cas des épidémies de virus SARS-Cov en 2003, les évènements de super-propagation étaient très fréquents, et on a estimé que la dispersion du nombre d'infections causée par chaque individu infecté avait un paramètre $k = 0.16$. Nous avons donc repris ce paramètre dans ce travail.

Nous avons utilisé les mêmes méthodes que précédemment pour calculer la vraisemblance mais en imposant cette fois une hétérogénéité des valeurs de $\mathcal{R}_0$ individuels.



Distribution des dates d'origine de la vague épidémique pour atteindre 100 décès le 20 mars avec hétérogénéité.

Comme attendu, nous constatons que cette surdispersion des cas secondaires augmente la vitesse de croissance des simulations où l'épidémie ne s'éteint pas toute seule. Par conséquent, la date d'origine de la vague épidémique estimée est plus tardive que dans le cas homogène. L'introduction des différences entre individus fait passer la date de début de la vague épidémique à médiane au **18 janvier** et 95 % des valeurs entre le 29 décembre et le 28 janvier.

Si l'origine de la vague épidémique est plus récente que dans le cas où les $\mathcal{R}_0$ individuels sont fixes, on constate tout de même que les intervalles de confiance se chevauchent. En augmentant l'hétérogénéité dans la population (donc en diminuant k), on pourrait parvenir à un différence significative mais cela se fait probablement au détriment du réalisme biologique.

L'augmentation de l'hétérogénéité dans la transmission du virus se traduit également par une augmentation de la variance dans les simulations, ce qui augmente l'incertitude quant à nos estimations de la date d'origine de la vague épidémique en France.

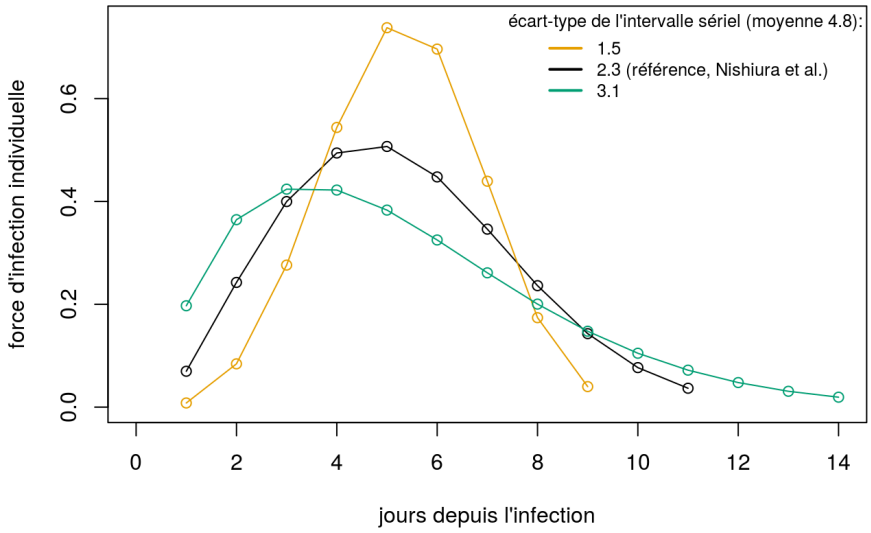
On peut, là encore, calculer la probabilité d'extinction d'une chaîne de transmission initiée par un individu. La seule différence par rapport au cas précédent réside dans l'hétérogénéité des valeurs de $\mathcal{R}_0$ individuels. Pourtant, cela suffit à faire bondir cette probabilité à 75 %. Intuitivement, ceci est lié au fait qu'en augmentant l'hétérogénéité, on va avoir beaucoup plus de super-propagateurs mais également beaucoup plus d'infections qui ne causent aucune infection secondaire. La probabilité d'extinction est directement dépendante de la proportion de ces nombres de reproduction individuels nuls.

Impact de l'intervalle sériel

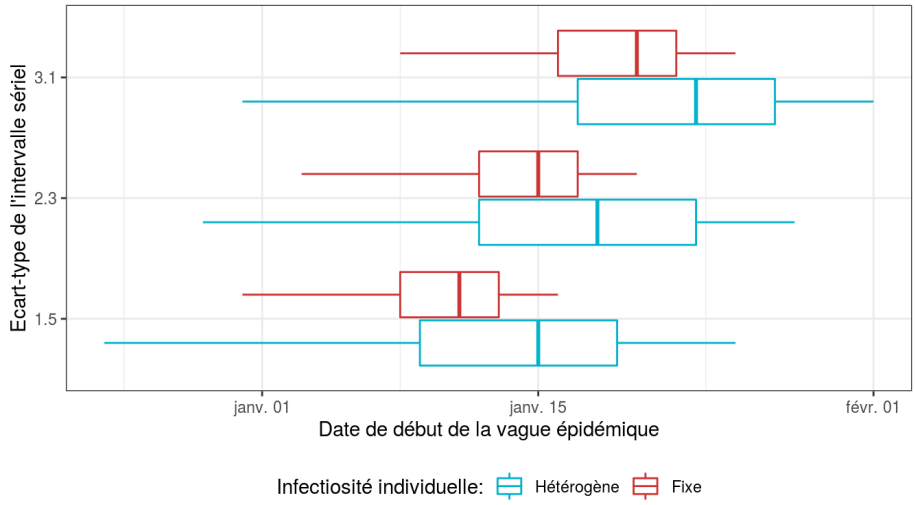
On l'a vu, un des paramètres clés de notre étude est l'*intervalle sériel*, qui est utilisé pour estimer le $\mathcal{R}_0$ et calculer la force d'infection individuelle. En l'absence de données disponibles en France, nous avons suivi Nishiura *et alii* (2020, *Int J Inf Dis* (https://linkinghub.elsevier.com/retrieve/pii/S1201971220301193)), qui estiment qu'en moyenne, lorsqu'un individu est infecté, il met 4,8 jours pour infecter à son tour une personne, avec un écart-type de 2,3 jours.

Quel est l'impact de ce choix sur le résultat ? Intuitivement, si la durée moyenne de l'intervalle sériel augmente, l'épidémie croissant moins vite, et faudrait donc remonter plus loin dans le temps pour trouver la date de début de la vague épidémique. Cependant, comme l'estimation du $\mathcal{R}_0$ dépend également de l'intervalle sériel, ces deux effets pourraient aussi se compenser.

Nous avons donc étudié l'effet de la dispersion de l'intervalle sériel sur la date de début de l'épidémie, en utilisant un écart-type de 1,5 jours, 2,3 jours (comme dans le reste de l'étude) ou 3,1 jours. Ces valeurs correspondent aux trois forces d'infection individuelles illustrées dans la figure suivante.



La figure suivante résume les effets de cette dispersion sur la date d'origine de la vague épidémique dans des populations homogènes (en rouge) ou hétérogènes (en bleu).



Les « boîtes à moustaches » sur le graphique représentent les quantiles : 2,5 %, 25 %, 50 %, 75 % et 97,5 %.

On observe que plus la dispersion de l'intervalle sériel est grande, plus la date de début de la vague épidémique est récente. La date médiane de début passe ainsi du 11 au 19 janvier dans le modèle où l'infectiosité de la population est homogène. Cependant cette variation reste au sein de la gamme de valeurs obtenues avec la distribution de l'intervalle sériel mesuré par Nishiura *et alii*.

Extinction de l'épidémie

Confinement prolongé

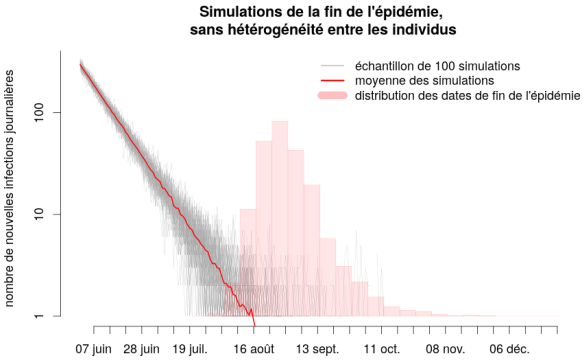
Nous allons maintenant explorer l'épidémie par son autre extrémité et tenter de déterminer combien de temps il aurait fallu prolonger le confinement pour être certains de ne plus avoir de cas dans la population.

Afin de limiter la stochasticité de la phase initiale de l'épidémie, nous utilisons les données issues d'un modèle détaillé dans le Rapport 3 (http://covid-ete.ouvaton.org/Rapport3_Modele.html) pour estimer le nombre de nouveaux infectés entre le 25 avril et le 06 mai et ainsi calibrer notre modèle.

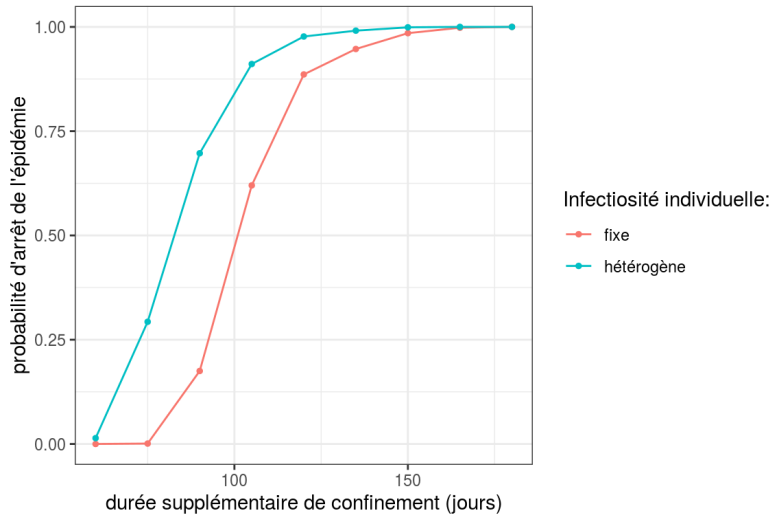
Pour ces simulations, nous avons conservé les même distributions de $\mathcal{R}_0$ que pour l'estimation de la date d'origine de l'épidémie et avons supposé que le confinement avait pour effet de diviser ce nombre de reproduction par environ 4 (soit une valeur moyenne de $\mathcal{R}_t = 0,66$). Le confinement a été levé au 11 mai et par conséquent depuis cette date le nombre de reproduction est supérieur. Toutefois, pour simplifier, nous allons ici supposer que cette levée n'a pas d'effet. Ceci peut être interprété comme une hypothèse très optimiste ou comme un exercice de pensée visant à voir combien de temps il aurait fallu prolonger le confinement afin de faire passer la taille de l'épidémie sous un certain seuil.

Notre but est de montrer le rôle de la stochasticité dans la disparition d'une épidémie. Pour cela, nous réalisons 1.000 simulations dans des conditions identiques et retenons, pour chaque simulation, la date à laquelle l'épidémie s'éteint.

Le premier graphique est un histogramme qui représente la distribution des dates d'extinction suivant l'hypothèse que la population est homogène. Les lignes grisées représentent les trajectoires épidémiques que l'on voit s'éteindre les unes après les autres au cours du temps. Attention, la légende correspond aux trajectoires épidémiques (nombre de nouvelles infections). On voit que même avec un confinement prolongé, il aurait fallu attendre en moyenne jusqu'à la mi-août pour ne plus avoir aucun cas de COVID-19 en France. Et encore, l'intervalle de confiance à 95 % inclut des dates bien plus tardives.



En supposant que la population est hétérogène (avec un paramètre $k = 0,16$), les résultats sont qualitativement similaires. Toutefois, afin de mieux visualiser la différence entre les deux, on peut représenter, pour une date donnée, le pourcentage des 1.000 simulations où l'épidémie s'est éteinte.



Sur la courbe rouge, on voit qu'avec 3 mois de confinement supplémentaire, il y avait moins d'une chance sur deux que l'épidémie se soit complètement éteinte. Et pour être sûr à plus de 90 % de ne plus avoir de cas, il aurait fallu attendre plus de 150 jours.

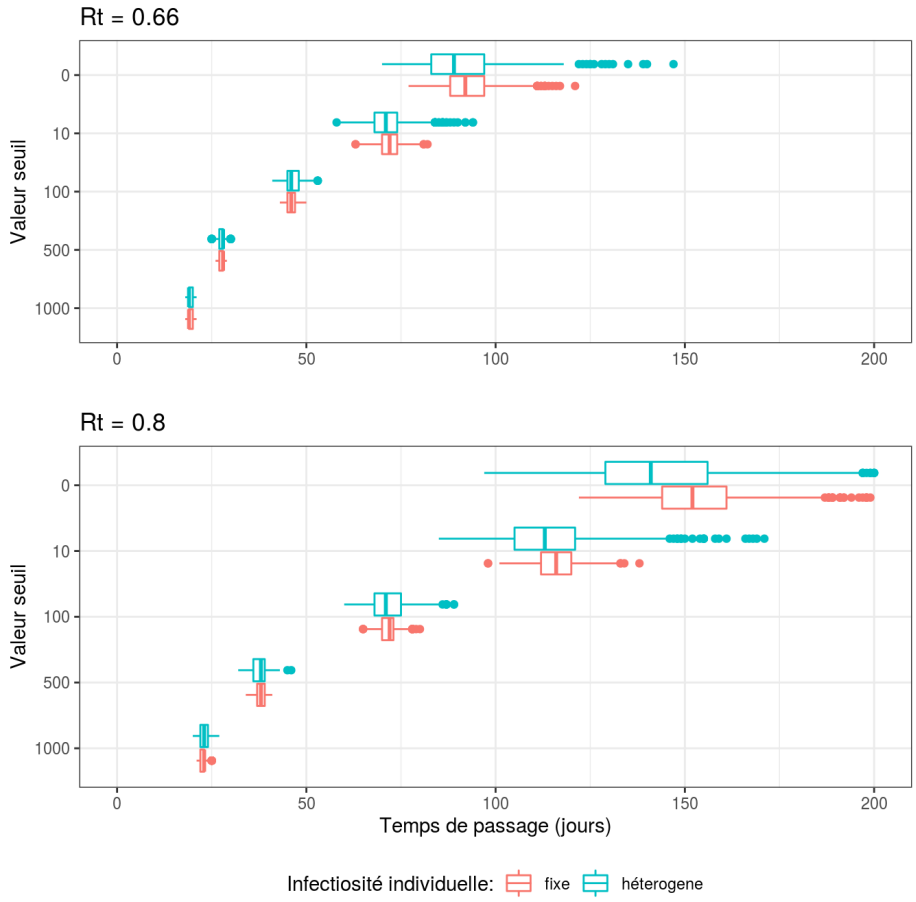
La courbe bleue du graphique représente un cas avec hétérogénéité individuelle. Comme attendu, ceci augmente la probabilité d'extinction de l'épidémie. Ainsi, avec 3 mois supplémentaires de confinement la probabilité que l'épidémie se soit complètement éteinte dépasse les 75 %.

Ces durées de confinement font un peu froid dans le dos mais il est possible de représenter le nombre de jours qu'il aurait fallu attendre pour atteindre des seuils plus limités. Par exemple, combien de jours pour être sûrs à 95 % d'avoir moins de 1.000 personnes infectées ? Ou 100 ? Ou 10 ? Nous représentons cela sur le tableau suivant, pour des cas avec ou sans hétérogénéité. Les valeurs entre crochets indiquent les intervalles de confiance.

Nombre maximal d'infections en cours	**Jours (R_0 fixé)**	**Jours (hétérogénéité des individus)**
1000	19 [Quantile 5%-95% : 19-20]	19 [Quantile 5%-95% : 18-20]
100	46 [Quantile 5%-95% : 44-48]	46 [Quantile 5%-95% : 43-50]
10	72 [Quantile 5%-95% : 67-77]	71 [Quantile 5%-95% : 64-81]
0	93 [Quantile 5%-95% : 83-105]	91 [Quantile 5%-95% : 76-110]

Relâchement

Pour conclure, nous représentons les nombre de jours à attendre pour que l'épidémie passe sous des valeurs seuils en termes de nombre de personnes infectées pour un cas plus réaliste ou après le 11 mai et la levée du confinement, le nombre de reproduction augmenterait à $\mathcal{R}_t = 0,8$. Afin de fournir une base de comparaison, nous représentons aussi le cas précédent correspondant à un confinement prolongé.



Comme on le voit, la levée du contrôle sur l'épidémie se traduit par une augmentation du nombre de jours à attendre avant sa disparition.

On constate également que l'effet de l'hétérogénéité de l'infectiosité entre les individus est essentiellement marqué lorsqu'on regarde la date d'extinction totale de l'épidémie, ce qui est le plus logique puisque c'est celle qui est la plus sensible aux évènements stochastiques. Enfin, cet effet est plus visible pour un $\mathcal{R}_0$ post-confinement valant 0.8, même si celui-ci reste encore modeste.

Discussion

Nous avons utilisé un modèle stochastique, c'est-à-dire explicitement aléatoire, afin de déterminer la date d'origine la plus vraisemblable de la vague épidémique.

Ce résultat est dépendant de la qualité des données utilisées (que nous ne contrôlons pas) et des hypothèses du modèle. Sur ce dernier point, plusieurs simplifications ou choix peuvent affecter la précision de nos estimations. Citons entre autres la proportion de létalité (Note_IFR.html) (fixée à 0,5 %), l'intervalle sériel inconnu en France, le nombre de reproduction de base (Note_R0.html) (fixé à $\mathcal{R}_0 = 3$), l'absence de structure en âge, ou encore l'absence de structure spatiale.

Avec toutes ces limitations, nous estimons qu'en l'état des connaissances, la vague épidémique aurait été initiée aux alentours du 15 janvier (avec un intervalle de confiance à 95 % entre le 5 et le 19 janvier). Si on rajoute dans le modèle de l'hétérogénéité entre les hôtes, cette date est un peu plus récente (le 18 janvier, avec 95 % des valeurs entre le 3 et le 27 janvier).

Ce dernier point illustre comment l'hétérogénéité entre les hôtes peut avoir un effet dans la propagation initiale de l'épidémie. En effet, la présence de super-propagateurs peut entraîner une croissance plus rapide du nombre de cas (la date de début de la vague épidémique serait alors plus tardive).

Une seconde source d'incertitude quant à la date de début de la vague épidémique est la distribution de l'intervalle sériel. Cependant, ces deux éléments ne modifient pas radicalement la date estimée, qui varie pour ses valeurs extrêmes entre le 24 décembre et le 31 janvier. Comme nous le décrivons dans notre note sur les origines de l'épidémie (Note_origine.html), ces ordres de grandeur sont cohérents avec nos estimations (Rapport6_BDSKY.html) obtenues à partir des données de séquences génétiques virales.

Il est à noter que la date que nous estimons n'est pas la date du premier cas de Covid-19 présent dans la population française, mais la date du premier cas à l'origine de la flambée épidémique. En effet, comme nous avons pu le mesurer, si nous prenons en compte de manière réaliste l'hétérogénéité de l'infectiosité des individus, 75% des infections n'entrainent pas de vague épidémique. Il est donc possible qu'il y ait eu des cas avant le début de la vague épidémique, mais qui n'ont pas eu suffisamment de contacts infectieux pour propager l'épidémie.

Nous avons ensuite utilisé ce même modèle afin d'étudier combien de temps il faudrait maintenir le taux de reproduction du virus au niveau de celui pendant le confinement. Cette seconde analyse ne dépend plus de l'estimation du taux de létalité ainsi que du délai entre l'infection et le décès, puisque nous regardons directement la date de la dernière infection. Cependant, ce modèle dépendant toujours des autres paramètres estimés (notamment l'intervalle sériel et le nombre de reproduction temporel $\mathcal{R}_t$) et reste très simpliste.

Même avec ce scénario "optimiste"théorique" dans lequel nous parvenons à maintenir la propagation du virus à un faible niveau, celui-ci risque de circuler encore pendant une longue durée avant de disparaître complètement : dans les cas les plus extrêmes, même avec une épidémie en décroissance, le virus peut circuler ainsi jusqu'en novembre 2020. Enfin, cette dernière date reste très hypothétique et dépend fortement de paramètres encore difficiles à estimer tels que l'efficacité des mesures barrière adoptées après le confienement. Une augmentation de 10% du taux de contact décale de près de deux mois nos estimations quant à la date de fin de l'épidémie.

Sources et remerciements

- Les auteurs remercient le calculateur à haute performance itrop (plateforme South Green (http://www.southgreen.fr)) de l'IRD de Montpellier pour la fourniture des ressource de calcul à haute performance, qui ont contribué aux résultats présentés dans ce travail (plus de détails sur bioinfo.ird.fr (https://bioinfo.ird.fr/)).
- L'équipe de modélisation de l'équipe ETE est composée de Samuel Alizon, Thomas Bénéteau, Marc Choisy, Gonché Danesh, Ramsès Djidjou-Demasse, Baptiste Elie, Yannis Michalakis, Bastien Reyné, Quentin Richard, Christian Selinger, Mircea T. Sofonea.
- L'équipe a reçu un soutien de la Région Occitanie dans le cadre de son appel « Urgence Recherche Covid-19 ».
- Thomas Bénéteau est financé par une bourse de la Ligue Contre le Cancer (https://www.ligue-cancer.net/).
- Contribution à ce travail :
    - conception du travail : toute l'équipe
    - réalisation des analyses : TB, BE, BR
    - rédaction du rapport : TB, BE, SA, MTS
    - approbation : toute l'équipe
    - contact : covid-ete@ouvaton.org (mailto:covid-ete@ouvaton.org)
    - site : covid-ete.ouvaton.org (http://covid-ete.ouvaton.org)
- (cc) BY-NC (http://creativecommons.org/licenses/by-nc/4.0/)
  Ce(tte) œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International (http://creativecommons.org/licenses/by-nc/4.0/).